P2.22B

USING SSM/I DATA AND COMPUTER VISION TO ESTIMATE TROPICAL CYCLONE INTENSITY

Richard L. Bankert*
Paul M. Tag
Naval Research Laboratory, Monterey,CA

1. INTRODUCTION

Satellite imagery and other remote sensing products often provide the only observational data of tropical cyclones. This is especially true in the western Pacific where aircraft reconnaissance missions stopped in 1987. Manual estimate procedures using satellite imagery (Dvorak, 1984) provide valuable assistance in determining tropical cyclone intensity. An objective Dvorak technique (Velden, et al., 1998) is currently being studied to enhance the manual method. In an effort to take advantage of the unique characteristics (Hawkins, et al., 1998) of Special Sensor Microwave/Imager (SSM/I) data, one Naval Research Laboratory effort (outside the scope of this paper) involves the computation of empirical orthogonal functions of SSM/I tropical cyclone data and presenting those values as inputs to a neural network to estimate the tropical cyclone intensity at a given imagery time (May, et al., 1997).

The algorithm applied in the research described here also uses SSM/I data, specifically the 85 GHz (H-pol) channel and a derived rain rate product. The 512x512 pixel imagery is cyclone-centered and image characteristics (computer vision features) are computed from the imagery data. A subset of these features is presented to a pattern recognition algorithm (k-nearest neighbor) and an intensity estimate is provided as output.

A description of the imagery characteristics (including available data and computer vision features) and feature selection methodology is provided in section two. Section three is a discussion of the algorithm used to automate the tropical cyclone intensity estimate and the current evaluation results. A summary follows in section four.

* *Corresponding author address:* Richard L. Bankert, Naval Research Laboratory, 7 Grace Hopper Ave, Monterey, CA 93943-5502;
e-mail: bankert@nrlmry.navy.mil

2. IMAGERY CHARACTERISTICS

For any supervised learning algorithm, "ground truth" data is needed to train and test the classifier or identification algorithm. To satisfy this requirement, 583 SSM/I images (512x512 pixels; cyclone-centered) and the associated best track intensity (in knots) are taken from 114 cyclones (from 1988 to 1997) in both the Atlantic (includes Caribbean Sea and Gulf of Mexico) and Pacific (eastern and western) basins. An example 85 GHz image is shown in Figure 1. To automate the tropical cyclone intensity estimation, characteristic features need to be computed. Using the 85 GHz (H-pol) channel and a derived rain rate product, 150 computer vision features are computed for each image. An example rain rate image is shown in Figure 2. See Figure 3 for a distribution of the 583 images according to intensity.

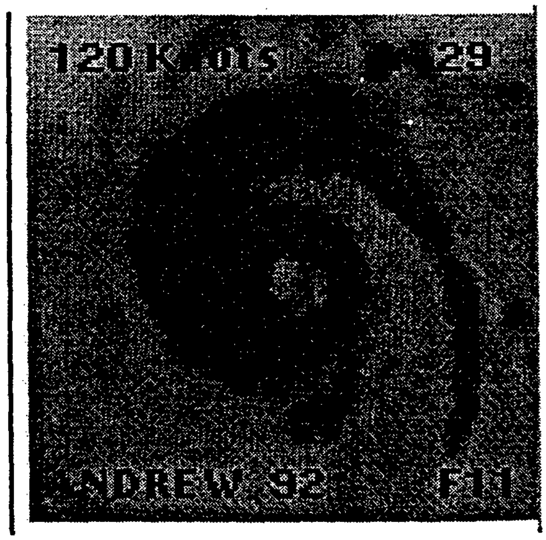

Figure 1. SSM/I 85 GHz (H-pol) image of Hurricane Andrew on 25 Aug 92 at 2252 UTC.

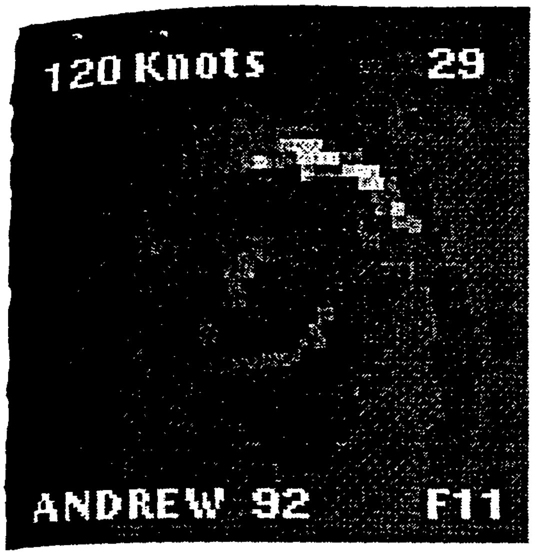

Figure 2. Derived rain rate image (using various SSM/I channels) of Hurricane Andrew on 25 Aug 92 at 2252 UTC (same as Figure 1).

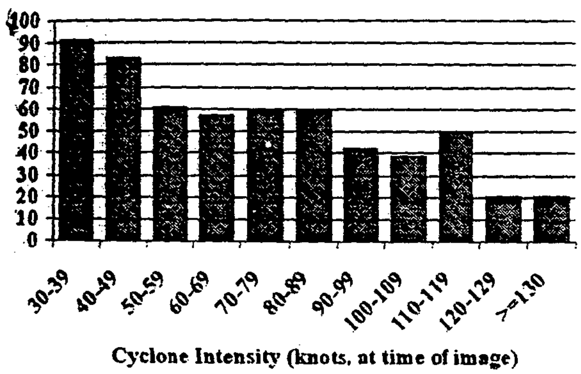

Figure 3. Distribution (by number) of the 583 training images, according to intensity.

The characteristic image features include computer vision attributes that can be described as size, shape, texture, and spectral measurements. Tropical cyclone-specific features include, among others, existence (yes/no) and size of an enclosed eye at a specific brightness temperature threshold (255K), the range of temperature thresholds for which an enclosed eye exists, and various rain rate measurements. Other features include latitude, longitude, date, time, etc.

Presenting redundant and irrelevant features to a pattern recognition algorithm will degrade its performance. To avoid this problem a feature selection algorithm is applied to the current data set. The selected feature subset is used as the input vector for the pattern recognition algorithm. For additional information on feature selection algorithms see Aha and Bankert (1995).

While previous uses of feature selection algorithms required the data to be organized in discrete classes, for this study the algorithm has been modified to make use of the actual intensity associated with each image (i.e, continuous data). Instead of searching for a feature subset that maximizes classfication accuracy, the search is for a subset that minimizes the root-mean-square error (measured in knots).

3. AUTOMATED INTENSITY ESTIMATE

A k-nearest neighbor classifier is used as the evaluation function in the feature selection algorithm and will serve as the automated tropical cyclone intensity estimate algorithm. This classification routine computes the similarity distances in feature space between the testing sample and each training sample. Using the single nearest neighbor distance as the standard for inclusion, those training samples within a distance factor (1.5 * nearest-neighbor distance) are used to estimate the testing sample intensity. A simple averaging technique is performed on those k-nearest neighbor intensities.

After a search and evaluation process within the feature selection algorithm, 18 characteristic features (listed in Table 1) were selected as the feature subset to represent each tropical cyclone at the time of the imagery. A leave-one-out cross validation test was applied to the entire data set. In this test, each sample (represented by the 18 selected features) is presented to the k-nearest neighbor algorithm, with the similarity distance to each of the remaining 582 samples computed. The feature selection and testing procedure is illustrated in Figure 4. The distribution of absolute error among 10 knot bins can be found in Figure 5; the average absolute error (AAE) for the 583 samples is 11.6 knots and the root mean square error (RMSE) is 15.8 knots.

Although the leave-one-out cross validation is considered a measure of the algorithm performance on unseen instances, two additional tests were performed on unseen cases. For each test, the data were divided into two parts: one to serve as the samples for feature selection and training and the other as testing. The first test was

a random selection (approx. 75% training, 25% testing). Twelve features were selected using 444 samples in the feature selection algorithm. Leave-one-out cross validation testing on these training samples resulted in an RMSE of 16.4 kts and an AAE of 12.5 kts (distance factor = 2.5). For the independent testing set (139 samples) the RMSE is 17.1 kts with an AAE of 13.5 kts (distance factor = 2.25).

Table 1. Selected features (583 samples).

1. # pixels > 0 rain rate
2. # TC (<255 K) pixels in NE quadrant
3. Max segmented region size / Total TC pixels
4. Latitude
5. Standard Dev. of TC pixels (inner 100x100)
6. Gray Level Diff. Vector contrast (N-S direction)
7. Sum and Diff. Hist. correlation (E-W direction)
8. Std. Dev. of quadrant sizes (inner 200x200)
9. Mean pixel value (inner 100x100)
10. Total rain rate (inner 150x150)
11. # pixels >0 rain rate (inner 150x150)
12. Median pixel value (> 255K pixels only)
13. # TC pixels (< 255K)
14. I Month - 9I
15. Minimum pixel value (inner 200x200)
16. Minimum pixel value (inner 100x100)
17. # TC pixels in SE quad / Total TC pixels
18. Highest enclosed eye threshold temperature - Lowest enclosed eye threshold temperature

The second test divided the samples by cyclone, with no images from a particular cyclone in both the training and testing sets. For example, all Andrew images are training samples and all Linda images are testing samples. Twelve features were selected using 439 training samples (103 tropical cyclones). Leave-one-out testing of these samples (distance factor = 1.5) resulted in an RMSE of 15.4 kts and AAE of 11.6 kts. For the independent testing set (144 samples; 11 tropical cyclones) the RMSE is 23.1 kts and the AAE is 18.5 kts (distance factor = 2.5).

All testing results are summarized in Table 2. Four of the five tests have very similar results. These results are very encouraging especially considering the fact that the best track intensity (used as ground truth) is recognized as being an approximation and not necessarily exact. The

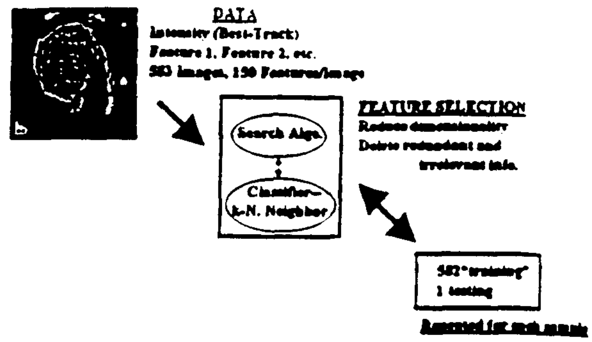

Figure 4. Procedural steps for extracting and selecting features with the imbedded leave-one-out cross validation test.

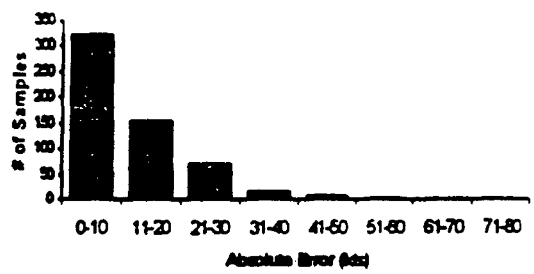

Figure 5. Distribution of absolute errors of estimated tropical cyclone intensity for 583 samples in a leave-one-out cross validation test.

Table 2. Automated tropical cyclone intensity estimate testing results. RMSE - root mean square error; AAE - average absolute error

TEST	*RMSE (kts)*	*AAE (kts)*
Leave-one-out (583 samples)	15.4	11.6
Leave-one-out (Random-444 samples)	16.4	12.5
Random Unseen (139 samples)	17.1	13.5
Leave-one-out (Cyclone-439 samples)	15.4	11.6
Cyclone Unseen (144 samples)	23.1	18.5

higher error statistics that resulted from the remaining test (cyclone unseen) can be explained in part to the fact that there were no images in the

ng set that were taken from the same cyclone the test sample. Therefore, there was no sibility of the k-nearest neighbor calculation g influenced by a training sample that is close time (and image characteristic features) to the ting sample. What this result clearly shows is at there are many variations in tropical cyclone aracteristics, as related to intensity, and that the aining database needs to be as large and as lusive as possible.

SUMMARY

An algorithm to estimate tropical cyclone intensity from SSM/I imagery using computer vision is demonstrated. A feature selection algorithm is applied to the data to reduce the dimension of the input vector and enhance the performance of the pattern recognition algorithm. A k-nearest neighbor algorithm is used to output a tropical cyclone intensity estimate at a particular SSM/I image time by calculating similarity distances (in the selected feature space) between the unknown sample and the stored training samples.

Testing results, while promising, indicate there is room for improvement. Adding more samples, examining other feature characteristics, and computing features from other SSM/I channels and derived products should improve performance. We will also examine ways to manipulate the k-nearest neighbor computations (distance factors, median instead of mean, etc.) to maximize performance on unseen samples.

5. ACKNOWLEDGMENT

The support of the sponsor, the Office of Naval Research, under Program Element 0602435N is gratefully acknowledged. The assistance of Jeff Hawkins (NRL Monterey) and Marla Helveston (Analysis and Technology) in supplying the data and offering technical advice is greatly appreciated. David Aha (NRL AI Center) is thanked for providing the feature selection algorithm and offering guidance.

6. REFERENCES

Aha, D.W., and R.L. Bankert, , 1995: A comparative evaluation of sequential feature selection algorithms. Proceedings of the Fifth International Workshop on AI and Statistics, Ft. Lauderdale, FL, 1-7.

Dvorak, V.F., 1984: Tropical cyclone intensity analysis using satellite data. NOAA Tech. Rep. NESDIS 11. U.S. Dept. of Commerce, 47 pp.

Hawkins, J.D., D.A. May, J. Sandidge, R. Holyer, and M.J. Heleveston, 1998: SSM/I-based tropical cyclone structural observations. Preprints 9th Conference on Satellite Meteorology and Oceanography, Amer. Meteor. Soc., 45 Beacon St., Boston, MA 02108, pages unknown.

May, D.A., J. Sandidge, R. Holyer, and J.D. Hawkins, 1997: SSM/I derived tropical cyclone intensities. Preprints 22nd Conference on Hurricanes and Tropical Meteorology, Amer. Meteor. Soc., 45 Beacon St., Boston, MA 02108, 27-28.

Velden, C.S., T.L. Olander, and R.M. Zehr, 1998: Development of an objective scheme to estimate tropical cyclone intensity from digital geostationary satellite infrared imagery. Accepted by *Weather and Forecasting*.